



PATENT ABSTRACTS OF JAPAN

(11) Publication number: **08314950 A**

(43) Date of publication of application: 29 . 11 . 96

(51) Int. Cl

G06F 17/30(21) Application number: **07114822**(71) Applicant: **FUJITSU LTD**

(22) Date of filing: 12 . 05 . 95

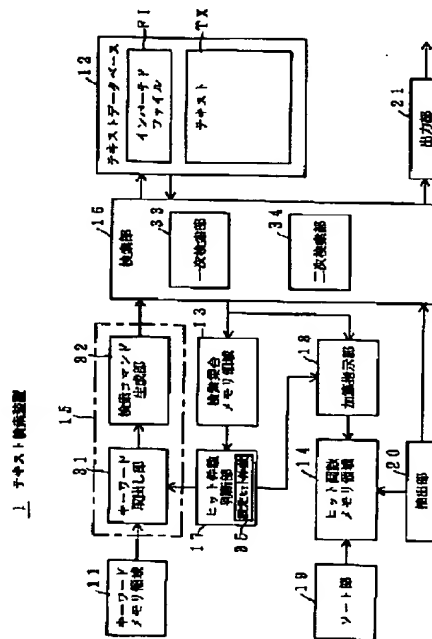
(72) Inventor: **TAKEYAMA YOSHIO**(54) **RETRIEVAL METHOD AND DEVICE FOR TEXT**

(57) Abstract:

PURPOSE: To systematically and comprehensively retrieve data through the use of a selected keyword and to easily and speedily extract a text fitted to a retrieval target.

CONSTITUTION: A keyword memory area 11 storing a keyword group, a retrieval instruction part 15 taking out the keyword one by one and instructing to perform retrieval, a primary retrieval part 33 retrieving the text, a retrieval set memory area 13 storing the text which hit, secondary retrieval part 34 executing secondary retrieval by AND obtained by adding another keyword when the number of hitting exceeds a set number of hitting, a hit number memory area 14 recording the number of hitting times of the text when the number of hitting times is less than the set number of hitting, an extraction part 20 extracting the text of the setting number of extraction in the order from one having the largest number of hitting times and outputting respective contents and an output part 21 are provided.

COPYRIGHT: (C)1996,JPO



(19) 日本国特許庁 (J P)

(12) 公開特許公報 (A)

(11) 特許出願公開番号

特開平8-314950

(43) 公開日 平成8年(1996)11月29日

(51) Int.Cl.⁶

G 0 6 F 17/30

識別記号

庁内整理番号

9194-5L

9194-5L

F I

G 0 6 F 15/403

15/40

3 7 0 Z

3 7 0 A

技術表示箇所

審査請求 未請求 請求項の数4 O L (全 8 頁)

(21) 出願番号 特願平7-114822

(22) 出願日 平成7年(1995)5月12日

(71) 出願人 000005223

富士通株式会社

神奈川県川崎市中原区上小田中4丁目1番
1号

(72) 発明者 竹山 善雄

東京都品川区東品川2-2-4 株式会社

富士通サポート・テクノロジー内

(74) 代理人 弁理士 久保 幸雄

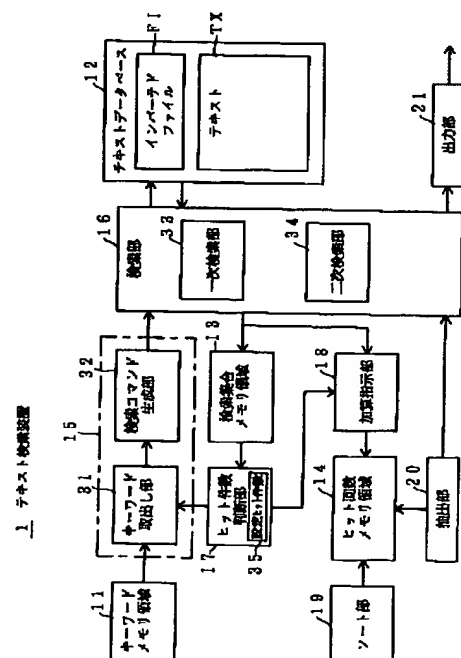
(54) 【発明の名称】 テキストの検索方法及び装置

(57) 【要約】

【目的】 選定されたキーワードを用いて系統的に且つ網羅的に検索を行い、検索目的に適合するテキストを容易且つ迅速に抽出することを目的とする。

【構成】 キーワード群を格納するキーワードメモリ領域11、1つずつキーワードを取り出して検索を指示する検索指示部15、テキストの検索を実行する一次検索部33、ヒットしたテキストを格納する検索集合メモリ領域13、ヒット件数が設定ヒット件数を越える場合に他の1つのキーワードを加えた論理積によって二次検索を実行する二次検索部34、ヒット件数が設定ヒット件数以下である場合にテキストのヒット回数を記録するヒット回数メモリ領域14、ヒット回数の多い順に設定抽出件数のテキストを抽出してそれぞれの内容を出力する抽出部20及び出力部21を有して構成される。

本発明に係るテキスト検索装置の構成を
概略的に示すブロック図



【特許請求の範囲】

【請求項 1】複数のテキストから構成されるテキストデータベースを備え、指定されたキーワード群によって前記テキストデータベースを検索する方法であって、

a) 指定されたキーワード群をキーワードメモリ領域内に格納し、

b) 前記キーワードメモリ領域から 1 つのキーワードを取り出し、

c) 取り出したキーワードを含むテキストを前記テキストデータベース内のテキスト群から検索し、ヒットしたテキストの集合である検索集合を検索集合メモリ領域内に格納し、

d) 検索によるヒット件数が予め設定された設定ヒット件数を越える場合に、前記キーワード群から次の 1 つのキーワードを取り出し、取り出したキーワードによって二次検索を行い、

e) ヒット件数が前記設定ヒット件数以下になるまで前記ステップ d を繰り返して実行し、

f) 検索によるヒット件数が予め設定された設定ヒット件数以下である場合に、

ヒットした各テキストについてのヒット回数をカウントし、

g) ステップ b 以降を繰り返して実行し、前記キーワード群の全てのキーワードを少なくとも 1 回用いて検索を行い、

h) ヒットによりカウントされたテキストの中から、ヒット回数の多い順に、予め設定された設定抽出件数に相当する件数のテキストを抽出し、抽出したテキストの内容を出力することを特徴とするテキストの検索方法。

【請求項 2】複数のテキストから構成されるテキストデータベースを備え、指定されたキーワード群によって前記テキストデータベースを検索する装置であって、指定されたキーワード群を格納するキーワードメモリ領域と、

前記キーワードメモリ領域から 1 つずつキーワードを取り出すとともに、格納されたキーワードについて少なくとも 1 回は取り出して検索を指示する検索指示手段と、キーワードによって前記テキストデータベースからテキストの検索を実行する一次検索手段と、

ヒットしたテキストの集合である検索集合を格納する検索集合メモリ領域と、

検索によるヒット件数が予め設定された設定ヒット件数を越えるか否かを判断する判断手段と、

検索によるヒット件数が予め設定された設定ヒット件数を越える場合に、前記キーワード群から取り出した他の 1 つのキーワードを加えた論理積によってテキストの検索を実行する二次検索手段と、

検索によるヒット件数が予め設定された設定ヒット件数以下である場合に、ヒットした各テキストについてのヒット回数を記録する記録手段と、

ヒットによりカウントされたテキストの中から、ヒット回数の多い順に、予め設定された設定抽出件数に相当する件数のテキストを抽出してそれぞれの内容を出力する出力手段と、

を有してなることを特徴とするテキストの検索装置。

【請求項 3】複数のテキストから構成されるテキストデータベースを備え、指定されたキーワード群によって前記テキストデータベースを検索する装置であって、

指定されたキーワード群を格納するキーワードメモリ領域と、

前記キーワードメモリ領域から 1 つずつキーワードを取り出すとともに、格納されたキーワードについて少なくとも 1 回は取り出して検索を指示する検索指示手段と、キーワードによって前記テキストデータベースからテキストの検索を実行する一次検索手段と、

ヒットしたテキストの集合である検索集合を格納する検索集合メモリ領域と、

検索によるヒット件数が予め設定された設定ヒット件数を越えるか否かを判断する判断手段と、

検索によるヒット件数が予め設定された設定ヒット件数を越える場合に、前記キーワード群から取り出した他の 1 つのキーワードを加えた論理積によってテキストの検索を実行する二次検索手段と、

テキストのヒット回数を記録するヒット回数メモリ領域と、

検索によるヒット件数が予め設定された設定ヒット件数以下である場合に、ヒットしたテキストについて前記ヒット回数メモリ領域のヒット回数に 1 を加算するよう指示する加算指示手段と、

前記ヒット回数メモリ領域に記録されたテキストの中から、ヒット回数の多い順に、予め設定された設定抽出件数に相当する件数のテキストを抽出する抽出手段と、抽出したテキストの内容を出力する内容出力手段と、を有してなることを特徴とするテキストの検索装置。

【請求項 4】前記検索指示手段は、

取り出されたキーワードが二次検索に用いられ且つ当該二次検索におけるヒット件数が零である場合に、当該キーワードを一次検索のためのキーワードとして再度用いるように指示する、

請求項 2 又は請求項 3 記載のテキストの検索装置。

【発明の詳細な説明】

【0001】

【産業上の利用分野】本発明は、日本文、英文などの複数のテキストの中から、指定されたキーワードを含むテキストを検索する方法及び装置に関する。

【0002】コンピュータのフィールドサポート業務において、新たに発生したトラブルの早期解決のために、過去において発生したトラブル事例をデータベースから検索してその解決方法を参照することが行われている。

すなわち、トラブルが発生した場合に、それが解決され

た後において、発生したトラブルの内容、それに対応してとられた処置、トラブルの原因、その他参考となる事項を文章で記述し、それを登録することによってトラブル事例のデータベースが構築されている。トラブルが新たに発生したときには、トラブルの内容に応じたキーワードを用いてデータベースを検索し、類似するトラブル事例を抽出してトラブル解決のために用いられている。

【0003】このようなシステムを実効あらしめるには、実際に参考となるトラブル事例を効率よく抽出することが必要である。

【0004】

【従来の技術】従来より、データベースから目的のテキストを抽出するに当たっては、利用者によってキーワードが選定され、選定されたキーワードがキーボードなどから入力された後、実行を指示することによって検索が行われる。検索に用いられたキーワードとそれによるヒット状況によって、目的のテキストが抽出される。検索目的に対しより適合するテキストを決定するために、さまざまなキーワードで検索が行われ、抽出されたテキストの内容が吟味される。複数のキーワードを選定することができる場合には、それらのキーワードの論理和又は論理積などによって検索が行われる。

【0005】一般に、高速のテキスト検索装置（文献検索システム、又は文章検索システムともいう）においては、データベースに蓄えられた大量のテキストの中から指定された検索条件を満足するテキストを高速で探索するためのインバーテッドファイルが設けられている（例えば、特開平3-294963号）。インバーテッドファイルの存在によって、ほとんどの検索条件がそのインバーテッドファイル上の操作だけで高速に処理されることとなる。

【0006】

【発明が解決しようとする課題】しかし、従来のテキスト検索装置によると、利用者が選定したキーワードによる検索と検索によって抽出されたテキストの吟味とを試行錯誤的に繰り返して検索が進められるので、網羅的な検索を行うことができず、検索漏れが生じたり、検索内容が片寄ったりするおそれがあった。

【0007】特に、指定されたキーワードの個数が多く、しかもそれらキーワードによる検索の途中においてヒット件数が多数にのぼることが分かった場合には、当初において予定していた検索の手順を大幅に変更してしまうことがあり、系統的な検索を行うのが実際には困難であるという問題があった。

【0008】本発明は、上述の問題に鑑みてなされたもので、選定されたキーワードを用いて系統的に且つ網羅的に検索を行い、検索目的に適合するテキストを容易且つ迅速に抽出することのできるテキストの検索方法及び装置を提供することを目的とする。

【0009】

【課題を解決するための手段】請求項1の発明に係る方法は、複数のテキストから構成されるテキストデータベースを備え、指定されたキーワード群によって前記テキストデータベースを検索する方法であって、

- a) 指定されたキーワード群をキーワードメモリ領域内に格納し、
- b) 前記キーワードメモリ領域から1つのキーワードを取り出し、
- c) 取り出したキーワードを含むテキストを前記テキストデータベース内のテキスト群から検索し、ヒットしたテキストの集合である検索集合を検索集合メモリ領域内に格納し、
- d) 検索によるヒット件数が予め設定された設定ヒット件数を越える場合に、前記キーワード群から次の1つのキーワードを取り出し、取り出したキーワードによって二次検索を行い、
- e) ヒット件数が前記設定ヒット件数以下になるまで前記ステップdを繰り返して実行し、
- f) 検索によるヒット件数が予め設定された設定ヒット件数以下である場合に、ヒットした各テキストについてのヒット回数をカウントし、
- g) ステップb以降を繰り返して実行し、前記キーワード群の全てのキーワードを少なくとも1回用いて検索を行い、
- h) ヒットによりカウントされたテキストの中から、ヒット回数の多い順に、予め設定された設定抽出件数に相当する件数のテキストを抽出し、抽出したテキストの内容を出力する。

【0010】請求項2の発明に係る装置は、複数のテキストから構成されるテキストデータベースを備え、指定されたキーワード群によって前記テキストデータベースを検索する装置であって、指定されたキーワード群を格納するキーワードメモリ領域と、前記キーワードメモリ領域から1つずつキーワードを取り出すとともに、格納されたキーワードについて少なくとも1回は取り出して検索を指示する検索指示手段と、キーワードによって前記テキストデータベースからテキストの検索を実行する一次検索手段と、ヒットしたテキストの集合である検索集合を格納する検索集合メモリ領域と、検索によるヒット件数が予め設定された設定ヒット件数を越えるか否かを判断する判断手段と、検索によるヒット件数が予め設定された設定ヒット件数を越える場合に、前記キーワード群から取り出した他の1つのキーワードを加えた論理積によってテキストの検索を実行する二次検索手段と、検索によるヒット件数が予め設定された設定ヒット件数以下である場合に、ヒットした各テキストについてのヒット回数を記録する記録手段と、ヒットによりカウントされたテキストの中から、ヒット回数の多い順に、予め設定された設定抽出件数に相当する件数のテキストを抽出してそれぞれの内容を出力する出力手段と、を有して

なる。

【0011】請求項3の発明に係る装置は、指定されたキーワード群を格納するキーワードメモリ領域と、前記キーワードメモリ領域から1つずつキーワードを取り出すとともに、格納されたキーワードについて少なくとも1回は取り出して検索を指示する検索指示手段と、キーワードによって前記テキストデータベースからテキストの検索を実行する一次検索手段と、ヒットしたテキストの集合である検索集合を格納する検索集合メモリ領域と、検索によるヒット件数が予め設定された設定ヒット件数を越えるか否かを判断する判断手段と、検索によるヒット件数が予め設定された設定ヒット件数を越える場合に、前記キーワード群から取り出した他の1つのキーワードを加えた論理積によってテキストの検索を実行する二次検索手段と、テキストのヒット回数を記録するヒット回数メモリ領域と、検索によるヒット件数が予め設定された設定ヒット件数以下である場合に、ヒットしたテキストについて前記ヒット回数メモリ領域のヒット回数に1を加算するよう指示する加算指示手段と、前記ヒット回数メモリ領域に記録されたテキストの中から、ヒット回数の多い順に、予め設定された設定抽出件数に相当する件数のテキストを抽出する抽出手段と、抽出したテキストの内容を出力する内容出力手段と、を有してなる。

【0012】請求項4の発明に係る装置では、前記検索指示手段は、取り出されたキーワードが二次検索に用いられ且つ当該二次検索におけるヒット件数が零である場合に、当該キーワードを一次検索のためのキーワードとして再度用いるように指示するように構成される。

【0013】

【作用】本発明の検索の手順の例を図8に示すフローチャートに基づいて説明する。指定されたキーワードは、キーワードメモリ領域11に格納される(#12)。キーワードメモリ領域11から1つのキーワードが取り出される(#13)。取り出したキーワードを用いて検索を行うためのコマンドが生成され(#14)、コマンドの実行によってテキストデータベース12の検索が行われる(#15)。

【0014】ヒットしたテキストTXが検索集合メモリ領域13に格納され(#16)、ヒット件数が設定ヒット件数以下である場合には(#17でイエス)、テキストTXがヒット回数メモリ領域14に登録され且つヒット件数に「1」が書き込まれる(#18)。

【0015】ヒット件数が設定ヒット件数を越える場合には(#17でノー)、ステップ#13に戻ってキーワードメモリ領域11から次のキーワードが取り出され、二次検索が行われる。

【0016】キーワードメモリ領域11に格納されたキーワードで1回も使用されていないキーワードがある場合には(#19でノー)、ステップ#13に戻ってキー

ワードメモリ領域11から次のキーワードが取り出され、一次検索が行われる。二次検索におけるヒット件数が「0」であった場合には、同一のキーワードによって一次検索が新規に行われる。

【0017】そして、キーワードメモリ領域11に格納された全てのキーワードについて使用した場合には(#19でイエス)、ヒット回数メモリ領域14に登録されたテキストTXがソートされ(#20)、そこから設定抽出件数に等しい件数の識別名IDが抽出され(#21)、その内容が出力される(#22)。

【0018】

【実施例】図1は本発明に係るテキスト検索装置1の構成を機能的に示すブロック図、図2はテキスト検索装置1のハード構成の例を示すブロック図、図3はテキストデータベース12の例を示す図である。

【0019】図1において、テキスト検索装置1は、キーワードメモリ領域11、テキストデータベース12、検索集合メモリ領域13、ヒット回数メモリ領域14、本発明の検索指示手段としての検索指示部15、検索部16、ヒット件数判断手段としてのヒット件数判断部17、記録手段及び加算指示手段としての加算指示部18、ソート部19、抽出手段としての抽出部20、出力手段及び内容出力手段としての出力部21などから構成されている。なお、ヒット回数メモリ領域14は本発明の記録手段としての機能をも有している。

【0020】ヒット件数判断部17には、制御定数である設定ヒット件数を格納する設定ヒット件数メモリ領域35が設けられている。キーワードメモリ領域11は、指定されたキーワード群を格納するメモリ領域である。図5に示すキーワードメモリ領域11には8個のキーワードKWが格納されている。キーワードKWの集合であるキーワード群は、図示しないキーボードその他の入力装置から入力される。又は、後で説明する実施例のように、入力された自然文CMからキーワードKWが自動的に抽出され、抽出されたキーワードKWがキーワードメモリ領域11に自動的に入力される。

【0021】テキストデータベース12は、インバーテッドファイルFI、及び多数のテキストTXのデータからなる。図3に示すように、各テキストTXはデータ長が不定である。テキストTXには、text1、text2…などの識別名IDが付されている。テキストデータベース12は、テキストTXの追加、変更、削除などの更新が可能である。インバーテッドファイルFIは、多数の索引語と各索引語を含むテキストTXの識別名IDとを対応付けて格納したものである。

【0022】検索集合メモリ領域13は、ヒットしたテキストTXの集合である検索集合を格納するメモリ領域である。ヒット回数メモリ領域14は、一次検索によって、又は二次検索によって、又は二次検索の繰返しによって、ヒット件数が設定ヒット件数以下になったテキ

ストTXの識別名IDを登録し、且つそのヒット回数を記録するメモリ領域である。

【0023】検索指示部15には、キーワード取出し部31、及び検索コマンド生成部32が設けられている。キーワード取出し部31によって、キーワードメモリ領域11から1つずつ順番にキーワードKWが取り出される。検索コマンド生成部32は、取り出したキーワードを用いて検索を行うためのコマンドを生成する。コマンドは、(命令語+キーワード)の形式で生成される。命令語として、一次検索を指示する「SEARCH」、及び二次検索を指示する「AND」がある。

【0024】キーワードメモリ領域11から最初に取り出したキーワードについては一次検索を指示する。一次検索のヒット件数が予め設定された設定ヒット件数を越える場合に、次に取り出されたキーワードについて二次検索を指示する。キーワードについて二次検索を指示し且つその二次検索におけるヒット件数が零であった場合には、当該キーワードによって一次検索を新規に指示する。

【0025】検索部16は、検索指示部15から受け取ったコマンドによって検索を実行する。検索部16には、一次検索手段としての一次検索部33、二次検索手段としての二次検索部34が設けられている。一次検索部33は、命令語が「SEARCH」であるときに、テキストデータベース12からキーワードによってテキストTXの一次検索を実行する。検索集合であるテキストTXの識別名IDは検索集合メモリ領域13に格納される。二次検索部34は、命令語が「AND」であるときに、検索集合メモリ領域13からキーワードによってテキストTXの二次検索を実行する。つまり、二次検索では、一次検索のキーワードと二次検索のキーワードとの論理積によるテキストTXの検索を行うこととなる。二次検索は、ヒット件数が設定ヒット件数以下になるまで繰り返される。二次検索による検索集合も検索集合メモリ領域13に格納される。

【0026】ヒット件数判断部17は、検索によるヒット件数が予め設定された設定ヒット件数を越えるか否かを判断する。つまり、ヒット件数判断部17は、検索部16による検索が実行される度毎にヒット件数をチェックする。

【0027】加算指示部18は、ヒット回数メモリ領域14に登録されたテキストTXについて、ヒット件数が設定ヒット件数以下になる度毎に、ヒット回数に1を加算するように指示する。但し、テキストTXが最初に登録されるときは「1」を設定する。

【0028】ソート部19は、ヒット回数メモリ領域14に登録されたテキストTXについて、ヒット回数の多い順に並べ替える。抽出部20は、ソート後のヒット回数メモリ領域14に登録されたテキストTXについて、予め設定された設定抽出件数と等しい件数のテキストT

Xを抽出する。これによって、ヒット回数メモリ領域14に登録されたテキストTXの中から、ヒット回数の多い順に、設定抽出件数のテキストTXが抽出される。

【0029】出力部21は、抽出部20によって抽出されたテキストTXについて、その内容をテキストデータベース12から抽出して外部機器に出力する。図2において、テキスト検索装置1のハードウェアは、処理装置41、記憶装置42、大容量記憶装置43、キーボード44、ディスプレイ装置45、及びプリンタ装置46などによって構成される。

【0030】記憶装置42には上述のテキストデータベース12が格納される。大容量記憶装置43には、大容量のテキストデータベース12が格納される。処理装置41は、キーワードメモリ領域11、検索集合メモリ領域13、ヒット回数メモリ領域14、及び設定ヒット件数メモリ領域35となるメモリ領域、テキストデータベース12の全部又は一部が転送されるメモリ領域、ワークエリアとして使用されるメモリ領域を有している。キーボード44は、利用者が検索を行うに当たってキーワードKWを入力するために用いられるものであり、また、検索を行う際、又はテキストTXのテキストデータベース12への登録又は更新の際に、そのテキストTX、コマンド、又はデータなどを入力するためにも用いられる。検索結果その他の情報は、ディスプレイ装置45に表示され又はプリンタ装置46により用紙に印刷される。また、通信回線によって他のコンピュータ又はファイル装置などに転送される。

【0031】次に、テキスト検索装置1の処理内容又は動作について説明する。図4は検索のための自然文CMの例を示す図、図5はキーワードメモリ領域11の内容の例を示す図、図6は検索の実行経緯を示す図、図7はヒット回数メモリ領域14の内容の例を示す図、図8は検索処理の手順を示すフローチャートである。なお、図に示す例において、設定ヒット件数は「12」、設定抽出件数は「9」とする。

【0032】まず、図4に示すような自然文CMを入力すると、自然文CMからキーワードKWが自動的に抽出され(#11)、図5に示すようにキーワードメモリ領域11に格納される(#12)。図5に示す例では、自然文CMが文字種毎に区切られ、連続する同一の文字種の単語がキーワードKWとなる。但し、ひらかなの単語はキーワードKWとしない。

【0033】キーワードメモリ領域11から1つのキーワードKWが取り出される(#13)。取り出したキーワードKWを用いて検索を行うためのコマンドが生成され(#14)、コマンドの実行によってテキストデータベース12の検索が行われる(#15)。

【0034】図に示す例では、最初にキーワード「JEF」が取り出され、コマンド「SEARCH JEF」が生成され、コマンド「SEARCH JEF」によ

て検索が行われる。

【0035】ヒットしたテキストTXの識別名IDが検索集合メモリ領域13に格納され(#16)、ヒット件数が設定ヒット件数以下である場合には(#17でイエス)、テキストTXの識別名IDがヒット回数メモリ領域14に登録され且つヒット件数に「1」が書き込まれる(#18)。

【0036】ヒット件数が設定ヒット件数を越える場合には(#17でノー)、ステップ#13に戻ってキーワードメモリ領域11から次のキーワードKWが取り出され、二次検索が行われる。

【0037】キーワードメモリ領域11に格納されたキーワードKWで1回も使用されていないキーワードKWがある場合には(#19でノー)、ステップ#13に戻ってキーワードメモリ領域11から次のキーワードKWが取り出され、一次検索が行われる。但し、二次検索におけるヒット件数が「0」であった場合には、同一のキーワードKWによって一次検索が新規に行われる。

【0038】図に示す例では、コマンド「SEARCH JEF」による一次検索によって1927件がヒットしたが(図6参照)、この件数は設定ヒット件数の「12」を越えているので、次のキーワードKWである「コード」によって二次検索が行われる(図5及び図6参照)。しかし、二次検索によるヒット件数は0件であるので、ヒット回数メモリ領域14には記録されない。また、ヒット件数が0件であるので、同一のキーワード「コード」によって一次検索が行われ、ヒット件数「1」が得られる(図6参照)。この1件のテキストTXがヒット回数メモリ領域14に記録される。次のキーワードKWである「6EA4」がキーワードメモリ領域11から取り出され、一次検索が行われる。ヒット件数が「14」であるので、次のキーワード「拡張」によって二次検索が行われ、ヒット件数「9」が得られる。9件のテキストTXがヒット回数メモリ領域14に記録される。

【0039】そして、キーワードメモリ領域11に格納された全てのキーワードKWについて使用した場合には(#19でイエス)、ヒット回数メモリ領域14に記録された識別名IDがソートされ(#20)、そこから設定抽出件数に等しい件数の識別名IDが抽出され(#21)、その内容が出力される(#22)。

【0040】図7に示す例では、ヒット回数メモリ領域14には、text4が3回、text1が3回、text2が3回、text5が2回というように記録されている。ヒット回数メモリ領域14に記録されたテキストTXの中から上位9件が抽出され、その内容がディスプレイ装置45に表示され、且つプリンタ装置46で印刷される。

【0041】上述の実施例によると、複数のキーワードKWの指示が行われると、それらキーワードKWの全てを用いてテキストデータベース12の検索が行われ、必

要な件数のテキストTXが抽出されその内容が出力される。しかも、キーワードKWに対するヒット件数が統計的に処理され、ヒット回数の多い順に自動的に抽出される。したがって、従来のように利用者のキーワード組み合わせ能力に依存することなく、網羅的な検索を行って目的のテキストTXを容易に且つ迅速に検索することができる。

【0042】また、自然文CMを入力することによってキーワードKWが自動的に生成されるので、キーワードKWの指定が容易である。キーワードKWの指定のためのキーワードKWの生成と、テキストデータベース12におけるインバーテッドファイルFIの作成のためのキーワードKWの抽出とを連動させ、互いに同一のキーワードKWが得られるようにしておくと、検索の元となる自然文CMと類似度合いの高いテキストTXを抽出することができる。

【0043】また、二次検索におけるヒット件数が0件であった場合に同一のキーワードKWによって一次検索を新規に行うので、そのキーワードKWが有効に用いられることとなり、検索漏れを少なくすることができる。

【0044】上述の実施例において、キーワードKWとして、テキストTX又は自然文CMに用いられた単語それ自体を用いたが、カタカナなどを全て英文字に変換することとし、キーワードKWとしては英文字を多く用いるようにしてもよい。二次検索を検索集合メモリ領域13に対して実行したが、一次検索のキーワードKWをも用いてテキストデータベース12に対して検索を行うことによって二次検索を実行してもよい。コマンドの形式などは適宜変更することができる。ヒット回数メモリ領域14に記録されたテキストTXをソートしたが、ソートすることなく、ヒット回数の順に抽出することとしてもよい。テキストTXの種類、内容、文字種、数量、キーワードKWの選定に方法、キーワードメモリ領域11からのキーワードKWの取り出し順序、その他、テキスト検索装置1の構成、処理内容、処理順序などは、本発明の主旨に沿って適宜変更することができる。

【0045】

【発明の効果】請求項1乃至請求項4の発明によると、選定されたキーワードを用いて系統的に且つ網羅的に検索を行い、検索目的に適合するテキストを容易に且つ迅速に抽出することができる。

【0046】請求項4の発明によると、キーワードが有効に用いられることとなり、検索漏れを少なくすることができる。

【図面の簡単な説明】

【図1】本発明に係るテキスト検索装置の構成を機能的に示すブロック図である。

【図2】テキスト検索装置のハード構成の例を示すブロック図である。

【図3】テキストデータベースの例を示す図である。

10

20

30

40

50

【図 4】 検索のための自然文の例を示す図である。

【図 5】 キーワードメモリ領域の内容の例を示す図である。

【図 6】 検索の実行経緯を示す図である。

【図 7】 ヒット回数メモリ領域の内容の例を示す図である。

【図 8】 検索処理の手順を示すフローチャートである。

【符号の説明】

1 テキスト検索装置

11 キーワードメモリ領域

* 12 テキストデータベース

13 検索集合メモリ領域

14 ヒット回数メモリ領域 (記録手段)

15 検索指示部 (検索指示手段)

17 ヒット件数判断部 (ヒット件数判断手段)

18 加算指示部 (記録手段、加算指示手段)

20 抽出部 (出力手段、抽出手段)

21 出力部 (出力手段、内容出力手段)

33 一次検索部 (一次検索手段)

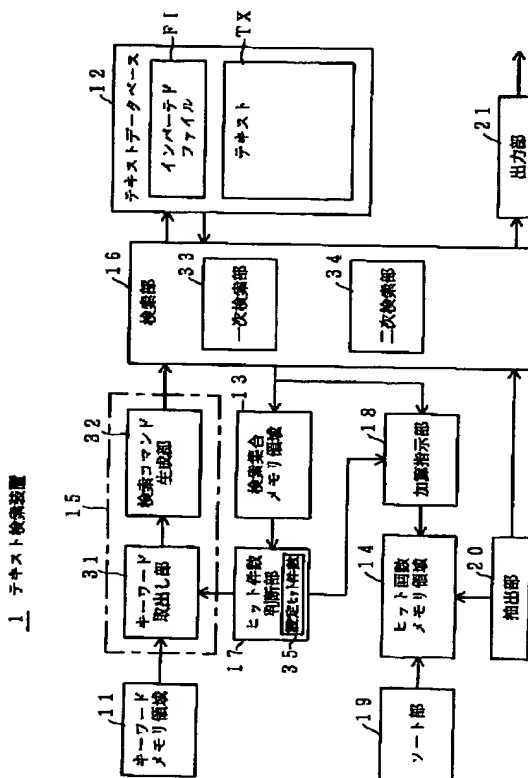
* 10 34 二次検索部 (二次検索手段)

【図 1】

【図 2】

【図 7】

本発明に係るテキスト検索装置の構成を機能的に示すブロック図



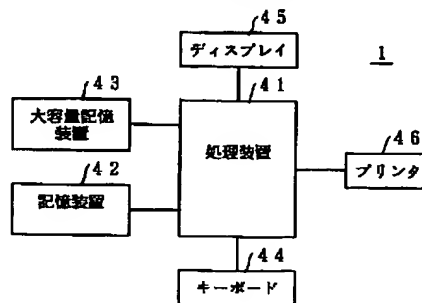
【図 3】

テキストデータベースの例を示す図

12 テキストデータベース

ID	TX
text 1	EP 3からXSPへ移行作業中。出力できな...
text 2	RDA-SVにて、JEF拡張漢字 X' 6...
text 3	NLPでJEFの6EA4の漢字と45FB...
text 4	JEF拡張漢字を含む文字セットを作成して...
text 5	拡張漢字をダイナミックローディングで、N...

テキスト検索装置のハード構成の例を示すブロック図



【図 4】

検索のための自然文の例を示す図

CM
JEFコード6EA4の拡張漢字が日本語プリンタに
印字できない

【図 5】

ヒット回数メモリ領域の内容の例を示す図

テキスト	ヒット回数
text 4	3
text 1	3
text 2	3
text 5	2
text 3	2

【図 6】

キーワードメモリ領域の内容の例を示す図

11 キーワードメモリ領域

RO	キーワード
1	JEF
2	コード
3	6EA4
4	拡張
5	漢字
6	日本語
7	プリンタ
8	印字

検索の実行経緯を示す図

	コマンド	ヒット件数
1	SEARCH JEF	1927
	AND コード	0000
2	SEARCH コード	0001
3	SEARCH 6EA4	0014
	AND 拡張	0008
4	SEARCH 漢字	6614
	AND 日本語	0048
	AND プリンタ	0000
5	SEARCH プリンタ	0008
6	SEARCH 印字	1529
	AND JEF	0057
	AND 6EA4	0002

【図 8】

検索処理の手順を示すフローチャート

